In press with J. Atmos. Sci., April 1999

Potential-Vorticity Inversion on a Hemisphere

Michael E. McIntyre and Warwick A. Norton*

Centre for Atmospheric Science at the

Department of Applied Mathematics and Theoretical Physics, Cambridge, UK

(http://www.atmos-dynamics.damtp.ac.uk/)

(Manuscript received June 1998, in final form April 1999)

* Present affiliation: Dept. of Atmospheric Oceanic and Planetary Physics, Oxford, UK.

Corresponding author address: M. E. McIntyre, Department of Applied Mathematics and Theoretical Physics, Cambridge CB3 9EW, UK

ABSTRACT

Several different kinds of accurate potential-vorticity (PV) inversion operators, and the associated balanced models, are tested for the shallow-water equations on a hemisphere in an attempt to approach the ultimate limitations of the balance, inversion, and slow-manifold concepts. The accuracies achieved are far higher than for standard balanced models accurate to one or two orders in Rossby number R or Froude number F (where $F = |\mathbf{u}|/c$, $|\mathbf{u}|$ = flow speed, c = gravity-wave speed). Numerical inversions, and corresponding balanced-model integrations testing cumulative accuracy, are carried out for cases that include substantial PV anomalies in the tropics. The balanced models in question are constructed so as to be exactly PV-conserving and to have unique velocity fields (implying, incidentally, that they cannot be Hamiltonian). Mean layer depths of 1 km and 2 km are tested.

The results show that, in the cases studied, the dynamical information contained in PV distributions is remarkably close to being complete even though $R = \infty$ at the equator and even though local maximum Froude numbers, F_{max}, approach unity in some cases. For example, in a 10-day integration of the balanced model corresponding to one of the most accurate inversion operators, "third order normal mode inversion", the mean depth was 1 km, the minimum depth less than 0.5 km, and $F_{\mathrm{max}} \simeq 0.7$, hardly small in comparison with unity. At the end of 10 days of integration, the cumulative r.m.s. error in the layer depth was less than 15 m, i.e., less than 5% of the typical r.m.s. spatial variation of 310 m. At the end of the first day of integration the r.m.s. error was 5 m, i.e., less than 2%. Here "error" refers to a comparison between the results of a balanced integration and those of a corresponding primitive-equation integration initialized to have low gravity-wave activity on day 0. Contour maps of the PV distributions remained almost indistinguishable by eye over the ten-day period. This remarkable cumulative accuracy, far beyond anything that could have been expected from standard scale analysis, is probably related to the weakness of the spontaneous-adjustment emission or "Lighthill radiation" studied in the companion paper (Ford et al. 1999).

1. Introduction

This paper presents the earliest results from a program begun in 1985 to investigate the concepts of balance, slow manifold, and potential vorticity (PV) inversion and to probe their accuracy and ultimate limitations. The theoretical and practical reasons for being interested in these concepts are now widely appreciated and need not be rehearsed here (e.g., Charney 1948, 1955, 1963, Gent and McWilliams 1984, McWilliams 1985, Hoskins et al. 1985, Lynch 1989, Allen 1993, Whitaker 1993, Warn et al. 1995, Vallis 1996, Grønås and Shapiro 1999, and refs., and many others). The dynamical system used is a hemispheric shallow-water model. Shallow-water systems, being the simplest continuum systems for which the concepts in question are nontrivial, are widely recognized as an important testbench for the associated theoretical ideas.

The results to be presented here were widely circulated in a preprint first submitted to *JAS* in 1990, based on the PhD thesis of Norton (1988), hereafter N88. A sample of the results, reproduced as Fig. 1 below, has already been published in connection with a discussion of zonally asymmetric wave–mean interaction phenomena (McIntyre and Norton 1990; for the wave–mean aspects see also the updates and corrections in Bühler and McIntyre 1998, and in Mo et al. 1998). However, full publication in the open literature has been delayed until now for a number of reasons, among which was an unresolved issue about local mass conservation arising from a discussion with J. S. Allen (personal communication). That issue was resolved in the course of recent work at the 1996 Isaac Newton Institute Programme on the Mathematics of Atmosphere and Ocean Dynamics, and will be briefly discussed in the final section of this paper along with related insights into Hamiltonian balanced models and Galilean invariance. Additional insights have become available, moreover, from the work of Ford et al. (1999, present issue of *JAS*), [[**EDITORS PLEASE CHECK**]], on the ultimate limitations of PV inversion imposed by the "Lighthill radiation" or spontaneous-adjustment emission of inertia–gravity waves by unsteady vortical flow (not to be confused with *Rossby adjustment*, in which the waves are generated by initial conditions). The work of Ford et al. applies to a parameter regime in which smallness of the Froude number F controls the error, as distinct from smallness of the Rossby number R. It presents an asymptotic analysis suggesting that, in the regime considered, the asymptotic order of accuracy reached here using what will be called "third order" PV inversion is the furthest one can go before encountering the ultimate limitations set by spontaneous-adjustment emission.

In terms of *numerical* accuracy the results presented here still provide, to our knowledge, the most accurate — and astonishingly accurate — nontrivial examples of balance and PV inversion yet found. In this regard we believe that we have pushed conditions closer to the ultimate limitations than in any work previously published, and with surprising success. The best accuracy attained is far greater, under the parameter conditions considered, than it would be for any of the standard balanced models such as quasigeostrophy, semigeostrophy, Bolin–Charney balance, or any other model based on scale analysis or asymptotic expansions correct to one or two orders in F or R when $F \ll 1$, $R \ll 1$.

The numerical accuracy tests included purely diagnostic tests like that summarized in Fig. 1, showing an example of what will be called "direct" third order PV inversion. The top pair of panels shows a set of surface-elevation, velocity, and divergence fields produced by primitive-equation evolution, and the middle pair shows the corresponding PV field in contours and in grayscale, with darker shades more cyclonic. The mean layer depth is 2 km. This is a complicated, unsteady, vortical flow with a vigorously meandering jet stream reaching well into the tropics. In the jet stream, local Froude numbers F, defined as the ratio of relative flow speed to local gravity wave speed, reach values $F_{\max} \simeq 0.5$. Rossby numbers $R \to \infty$ as we approach the equator. Such values can hardly be considered small in comparison with unity. The bottom pair of panels shows the surface elevation, velocity, and divergence fields reconstructed by PV inversion, i.e., using only the information represented by the middle pair of panels. To see the differences, even in the divergence fields contoured on the right, one has to look carefully.

Going beyond purely diagnostic tests, we also present prognostic tests of cumulative error over a few eddy turnaround times, at even larger F. These still more stringent tests are from ten-day runs of the PV-conserving balanced models associated with several different inversion operators, for a similar meandering-jet flow but with mean layer depth 1 km. F reaches a maximum value $F_{\max} > 0.7$ in the jet stream. When the most accurate inversion operators were used, contour maps of the PV distributions remained almost indistinguishable by eye over the ten-day period (e.g., Figs. 5b and 7f below). It is this, above all, that justifies the phrase "astonishingly accurate". We are dealing with an unsteady flow exhibiting nonlinear vortex interactions: it is sensitive to initial conditions, i.e., is of limited predictability, and presumably chaotic in the dynamical-systems sense.

The plan of the paper is as follows. Section 2 briefly summarizes the approach to be taken. Sections 3 and 5 present and illustrate two hierarchies of practically realizable PV inversion operators on the hemisphere, to be referred to as "direct" and "normal mode" inversion operators; sections 4 and 6–7 present numerical results, including the abovementioned cumulative accuracy tests (section 6). Section 7 shows how the PV inversion operators can be made Galilean invariant in an appropriate sense, and section 8 presents concluding remarks on conservation and related issues.

The present paper, in its most important respects, has only minor changes from the original version circulated in preprint form, except that most of the original background discussion, which now seems excessively lengthy — though needing no substantial revision — has been deleted along with details for less interesting cases with mean depth 8 km. Also deleted is part of the original section on Galilean invariance where the mathematical formulation was in error, as kindly pointed out to us by M. J. P. Cullen (personal communication, 1996). Some of the original background discussion has, however, been incorporated into the companion paper by Ford et al. (1999). One last but important point: while this paper was being reviewed, a error was discovered in the codes that implemented second and third order direct inversion numerically. We are grateful to Dr Eduard Neven for his close scrutiny of the codes, leading to discovery of the error, which is described in Appendix A. Tests with the corrected codes show clearly, however, that this error produces no significant differences in the results. For instance, in divergence difference maps like that in figure 3d below, produced with and without the coding error, the contours shifted by no more than about the plotted contour thickness, even with the minimal mean depth of 1 km and despite the sensitive nature of divergence difference maps. More precisely, the changes in such maps due to the coding error were of the order of 0.06 of a single contour interval. Such changes are inconsequential for the purposes of this paper.

2. Approach

Attention is restricted first of all to PV inversion operators for which the associated "PV-conserving balanced models" are not only as accurate as we can straightforwardly make them, but which also have the conceptually simplest structure placing no *a priori* limitations on closeness to primitive-equation behavior. We decided at the outset to avoid the problems of high-order filtered equations based on across-the-board expansion methods,[1]

[1] *Note added in 1998:* The problems associated with (asymptotic) expansion methods, as normally used to derive filtered equations for $F \ll 1$ or $R \ll 1$, have been discussed very carefully by Warn et al. (1995). Their paper points out that the most serious source of trouble is the use of expansions to approximate the PV evolution equation, or other *prognostic* equation or equations governing or influencing the 'slow' or vortical evolution — as distinct from the use of expansions to construct PV inversion operators while leaving the PV evolution equation unexpanded. Thus their approach is, as they acknowledge, similar to ours in its most essential aspects. See also the related discussions in Allen (1993), Vallis (1996), and in the present companion paper (Ford et al. 1999).

and in particular to retain the exact definition and evolution equation for the PV.

The PV-conserving balanced model associated with a given inversion operator is defined, then, simply to mean the dynamical system that results from timestepping the exact PV evolution equation, then inverting to get the new velocity field, then timestepping again, and so on, the limit of small timesteps being understood. This is evidently a balanced or filtered model in the usual sense that its prognostic equation is first order in time (e.g., Lynch 1989, & refs.). Within the complete set of equations, there is just one time derivative acting on just one scalar field. Such a model excludes, by definition, the possibility of any "slow-quasi-manifold fuzziness" related to motions resembling gravity waves (Warn 1997, Ford et al. 1999, & many refs.). From here onwards the term "gravity wave" will be defined to include all freely propagating inertia–gravity waves and tropical Kelvin waves. In the case of exactly conservative dynamics,[2]

[2] "Conservative dynamics" here means motion under conservative forces, i.e., the pressure-gradient, gravitational-centrifugal, and Coriolis forces, as distinct from "nonconservative forces" like boundary layer friction, parameterized gravity wave drag, and any other apparent force arising from the averaged effects of unresolved small-scale processes. Cf. footnote 3.

the prognostic equation is taken to be simply

$$DQ/Dt = 0 \ , \tag{2.1}$$

expressing exact *material* conservation of PV (Rossby 1936), with appropriate generalizations in the case of nonconservative dynamics, where Q is Rossby's exact shallow-water PV, Eq.(3.1d) below. The "nonconservative" generalizations will be left aside here, except to note in passing that they also express the idea of PV conservation in the *general* sense of that term.[3]

[3] This is the sense involving nonadvective fluxes, used for instance in connection with conservation of energy, momentum, and chemical tracers. For the corresponding properties of PV — more precisely, of "PV-substance" or "charge" — whose amount per unit mass is the PV, by definition, and which has no interior sources or sinks — the interested reader may consult Haynes and McIntyre (1990, & refs.). These properties justify, in particular, the unqualified use of the term "PV-conserving," even though it is the case of material conservation, as in Eq. (2.1), that we are specifically concerned with in this paper. It should be added though that balance and PV inversion can themselves be affected by, for instance, the mean effects of gravity waves (Bühler and McIntyre 1998).

Throughout this paper, then, we exclude from consideration any balanced or filtered model with a prognostic equation whose form differs from that of the exact PV equation, used with the exact PV formula. We require furthermore that a given PV-conserving balanced model has a unique velocity field, $\mathbf{u}(\mathbf{x}, t)$ say (unlike, for instance, quasigeostrophic or semigeostrophic theory), so that each of the preceding requirements has an unambiguous meaning. Here $\mathbf{x}$ is horizontal position and t is time. It is thus possible to speak unambiguously of *the* velocity field that results from applying a given inversion operator, and *the* advection of Q; and there is just one natural way of judging the accuracy of the resulting balance and PV inversion concepts, namely in terms of comparisons between primitive-equation behavior and the behavior of the PV-conserving balanced model in question.[4]

[4] For this purpose it is understood, of course, that the primitive equations are initialized to have sufficiently small gravity-wave activity, since we are not interested in gravity waves produced by Rossby adjustment, i.e., waves whose generation is attributable to imbalance in the initial conditions. We are grateful to a reviewer for pointing out that this should be kept in mind.

3. Direct inversion

We begin with a hierarchy of PV inversion operators to be referred to as "direct" inversion operators. For expository purposes we first describe them for the case of an unbounded f-plane; details for the hemispherical case are set out in Appendix A. To motivate the definitions we shall need to refer to the primitive equations. For our shallow water system, a single layer of fluid of depth $h(\mathbf{x}, t)$ on an f-plane, the primitive equations are conveniently taken as the vorticity, divergence and mass-conservation equations in the form

$$\frac{\partial \zeta}{\partial t} + f\delta = -\nabla.(\mathbf{u}\zeta) \ , \tag{3.1a}$$

$$\frac{\partial \delta}{\partial t} + \nabla^2 \phi' - f\zeta = -\nabla.(\mathbf{u}.\nabla\mathbf{u}) \ , \tag{3.1b}$$

$$\frac{\partial \phi'}{\partial t} + \hat{\phi}\delta = -\nabla \cdot (\mathbf{u}\phi') \ , \tag{3.1c}$$

assuming no bottom topography. Here ϕ' is the departure of the geopotential $\phi = gh$ from a constant reference value $\hat{\phi}$, g is the gravity acceleration, $\mathbf{u}$ is the horizontal velocity as before, ζ is the vertical component of relative vorticity, δ is the divergence, and f is the Coriolis parameter, equal to $2\Omega \sin\theta$, where θ is latitude. The shallow-water PV,

$$Q \ = \ \frac{f+\zeta}{h} \ = \ g\,\frac{f+\zeta}{\hat{\phi}+\phi'} \ , \tag{3.1d}$$

is an exact material invariant of these equations, as first shown by Rossby (1936).

It will prove convenient to introduce some special notation. Let $\mathrm{curl}^{-1}\zeta$ denote the rotational part of the velocity field $\mathbf{u}$ and $\mathrm{div}^{-1}\delta$ the divergent part, defined in the standard way through Helmholtz decomposition as

$$\mathrm{curl}^{-1}\delta \ = \ \mathbf{k}\times\nabla\psi \qquad \text{and} \qquad \mathrm{div}^{-1}\delta \ = \ \nabla\chi \ ,$$

with $\mathbf{k}$ a unit vertical vector. The streamfunction ψ and velocity potential χ satisfy

$$\nabla^2\psi = \zeta \qquad \text{and} \qquad \nabla^2\chi = \delta \ ,$$

together with suitable boundary conditions, here evanescence at infinity or, in Appendix A, suitable cross-equatorial symmetry conditions on the hemisphere.

The first and least accurate member of the hierarchy to be considered will be called "first order direct inversion". It is obtained, following the ideas of Bolin (1955), Charney (1955, 1962), and others, by deleting the time derivative in the divergence equation (3.1b) to give

$$\nabla^2\phi' - f\zeta = -\nabla\cdot(\mathbf{u}\cdot\nabla\mathbf{u}) \ . \tag{3.2a}$$

However, at this order we make $\mathbf{u}$ nondivergent by setting

$$\mathbf{u} = \mathrm{curl}^{-1}\zeta \ . \tag{3.2b}$$

The motivation is that when gravity wave activity is minimal, with one or both of the Froude and Rossby numbers small, $F \ll 1$ and $R \ll 1$, the term $\partial\delta/\partial t$ in (3.1b) tends to be small of order $\min(F^2, R^2)$ relative to typical magnitudes of the largest remaining terms, and the relative error in (3.2b) small of order $\delta/\zeta \simeq \min\{\max(F^2, F^2/R), R\}$. The ideas are standard; some relevant scale analysis may be found for instance in Haltiner and Williams (1980) and in McWilliams (1985). The formula (3.1d) defining the PV can be rearranged without approximation as

$$\zeta - \frac{f\phi'}{\hat{\phi}} = \frac{Q'}{g}(\hat{\phi} + \phi') \ , \tag{3.2c}$$

where Q' is the PV anomaly, defined as the departure of Q from its reference value $\hat{Q} = f/\hat{h}$, and where $\hat{h} = g^{-1}\hat{\phi}$, the area-averaged depth of the fluid layer.

Eqs. (3.2a–c) define the first order direct *inversion operator*, and Eqs. (3.2a–b) the corresponding *balance condition*. When the PV is given, (3.2a–c) make up a complete set of equations for the unknowns $\phi', \mathbf{u}, \zeta$. If we append to (3.2a–c) a single prognostic equation for DQ/Dt, such as $DQ/Dt = 0$, then we have the corresponding *PV-conserving balanced model.*

The second member of the hierarchy, to be called second order direct inversion, is the lowest member to yield a velocity field that has nonvanishing divergence δ. It is convenient to introduce auxiliary fields $\mathbf{u}_1$ and ζ_1 to be solved for as part of the inversion operation. They are diagnostic estimates of the corresponding time derivatives $\partial\mathbf{u}/\partial t$ and $\partial\zeta/\partial t$, in a sense to be made precise shortly; $\mathbf{u}_1$, for instance, has the dimensions of acceleration. We define the second order inversion operator by the following closed set of equations for the unknowns $\phi', \delta, \mathbf{u}, \mathbf{u}_1, \zeta_1, \zeta$:

$$\nabla^2\phi' - f\zeta = -\nabla.(\mathbf{u}.\nabla\mathbf{u}) , \tag{3.3a}$$

$$\mathcal{L}\,\delta = \nabla.\left\{f\zeta\mathbf{u} + \mathbf{u}_1.\nabla\mathbf{u} + \mathbf{u}.\nabla\mathbf{u}_1 - \nabla^2(\phi'\mathbf{u})\right\} , \tag{3.3b}$$

$$\mathbf{u} = \mathrm{curl}^{-1}\zeta + \mathrm{div}^{-1}\delta , \tag{3.3c}$$

$$\mathbf{u}_1 = \mathrm{curl}^{-1}\zeta_1 , \tag{3.3d}$$

$$\zeta_1 + f\delta = -\nabla.(\mathbf{u}\zeta) , \tag{3.3e}$$

$$\zeta - \frac{f\phi'}{\hat\phi} = \frac{Q'}{g}(\hat\phi + \phi') , \tag{3.3f}$$

where $\mathcal{L}$ is the linear omega-equation or modified Helmholtz operator,

$$\mathcal{L} = \hat\phi\nabla^2 - f^2, \tag{3.4}$$

whose inverse is robustly well behaved under suitable boundary conditions. The latter can be taken again as evanescence at infinity, or cross-equatorial symmetry/antisymmetry conditions on the hemisphere, as appropriate. Note that Eq. (3.3a) has the same appearance as (3.2a), but that $\mathbf{u}$ has a different meaning because (3.3c) includes a divergent part, unlike (3.2b). To motivate the remaining equations and the introduction of the auxiliary fields $\mathbf{u}_1, \zeta_1$, first compare (3.3b) with the generalized wave equation

$$\frac{\partial^2\delta}{\partial t^2} - \mathcal{L}\delta = -\nabla.\left(f\zeta\mathbf{u} + \frac{\partial}{\partial t}(\mathbf{u}.\nabla\mathbf{u}) - \nabla^2(\phi'\mathbf{u})\right) . \tag{3.5}$$

This is an exact consequence of the primitive equations (3.1), obtained by substituting from (3.1a) and ∇^2(3.1c) into ∂(3.1b)$/\partial t$. We expect the first term of (3.5) to be relatively small when gravity wave activity is minimal; formally it is of relative order $\min(F^2, R^2)$ as before. To get (3.3b) we delete this term and replace the time derivative $\partial \mathbf{u}/\partial t$ on the right by the auxilary field $\mathbf{u}_1$.

The latter field, which plays an important role in the tropics, can be thought of as an approximate diagnostic estimate of $\partial \mathbf{u}/\partial t$ that is to be sharply distinguished not only from the exact $\partial \mathbf{u}/\partial t$ of primitive equation evolution, but also from the approximate $\partial \mathbf{u}/\partial t$ that would be calculated if one were to timestep the PV-conserving balanced model consisting of a prognostic equation like $DQ/Dt = 0$ together with Eqs. (3.3).

Note also that Eq. (3.3e) has the same form as the exact vorticity equation except that $\partial\zeta/\partial t$ is replaced by the auxiliary field ζ_1, a diagnostic estimate of $\partial\zeta/\partial t$, which corresponds to $\mathbf{u}_1$ through (3.3d).

We may summarize the foregoing by saying that Eqs. (3.3) are obtained from (3.1a–c) and ∂(3.1b)$/\partial t$ by deleting the terms $\partial\delta/\partial t$ and $\partial^2\delta/\partial t^2$ and then treating the remaining time derivatives $\partial \mathbf{u}/\partial t$ and $\partial\zeta/\partial t$ as independent diagnostic variables, the auxiliary fields $\mathbf{u}_1$ and ζ_1. This gives us a closed, purely diagnostic problem, as required by the PV-inversion concept.

The foregoing is closely related to ideas put forward by Hinkelmann (1969) in the context of forecast initialization. Related ideas have been used in numerical weather forecasting under the label "implicit nonlinear normal mode initialization" (Temperton 1988, 1989 & refs.). Again, we may interdistinguish (a) the second order direct inversion operator defined by Eqs. (3.3a–f), (b) the corresponding balance condition (3.3a–e), and (c) the corresponding PV-conserving balanced model consisting of (3.3a–f) together with the prognostic equation for DQ/Dt.

The third member of the hierarchy, to be called third order direct inversion, follows a similar pattern. We take (3.1a–c), ∂(3.1b)$/\partial t$, and ∂^2(3.1b)$/\partial t^2$, delete $\partial^2\delta/\partial t^2$ and $\partial^3\delta/\partial t^3$, and treat the remaining time derivatives as independent diagnostic variables.

These are distinguished by numerical suffixes as before; thus for instance $\mathbf{u}_2$ is a diagnostic estimate of $\partial^2\mathbf{u}/\partial t^2$:

$$\nabla^2\phi' - f\zeta = -\nabla.(\mathbf{u}.\nabla\mathbf{u}) - \delta_1 \ , \tag{3.6a}$$

$$\mathcal{L}\,\delta = \nabla.\Big\{f\zeta\mathbf{u} + \mathbf{u}_1.\nabla\mathbf{u} + \mathbf{u}.\nabla\mathbf{u}_1 - \nabla^2(\phi'\mathbf{u})\Big\} \ , \tag{3.6b}$$

$$\begin{aligned}\mathcal{L}\,\delta_1 = \nabla.\Big\{&f\zeta_1\mathbf{u} + f\zeta\mathbf{u}_1 + \mathbf{u}_2.\nabla\mathbf{u} + 2\mathbf{u}_1.\nabla\mathbf{u}_1 + \mathbf{u}.\nabla\mathbf{u}_2 \\ &- \nabla^2(\phi'_1\mathbf{u} + \phi'\mathbf{u}_1)\Big\} \ ,\end{aligned} \tag{3.6c}$$

$$\mathbf{u} = \mathrm{curl}^{-1}\zeta + \mathrm{div}^{-1}\delta \ , \tag{3.6d}$$

$$\mathbf{u}_1 = \mathrm{curl}^{-1}\zeta_1 + \mathrm{div}^{-1}\delta_1 \ , \tag{3.6e}$$

$$\mathbf{u}_2 = \mathrm{curl}^{-1}\zeta_2 \ , \tag{3.6f}$$

$$\zeta_1 + f\delta = -\nabla.(\mathbf{u}\zeta) \ , \tag{3.6g}$$

$$\zeta_2 + f\delta_1 = -\nabla.(\mathbf{u}_1\zeta + \mathbf{u}\zeta_1) \ , \tag{3.6h}$$

$$\phi'_1 + \hat{\phi}\delta = -\nabla.(\mathbf{u}\phi') \ , \tag{3.6i}$$

$$\zeta - \frac{f\phi'}{\hat{\phi}} = \frac{Q'}{g}(\hat{\phi} + \phi') \ . \tag{3.6j}$$

It is this inversion operator, or rather its counterpart on the hemisphere (Appendix A), that produced the bottom pair of panels in Fig. 1. The corresponding balance condition is defined by (3.6a)-(3.6i). Note that (3.6a) is the same as the *exact* divergence equation (3.1b) except for the replacement of $\partial\delta/\partial t$ by δ_1. Eq. (3.6i) similarly corresponds to the exact mass-conservation equation (3.1c), and (3.6h) to the time derivative of the exact vorticity equation (3.1a).

The sequence of direct inversion operators can be extended to higher orders following precisely the same pattern, deleting successive pairs of time derivatives. All such operators have the usual "sign-reversal property", as discussed in Ford et al. (1999). This corresponds to the sign-reversal property of the PV itself, i.e., to the fact that if we keep the mass distribution unchanged while reversing the sign of the absolute velocity field, then the PV field reverses sign everywhere. Thus each inversion operator defined above is invariant under the transformation

$$Q, f, \mathbf{u}, \delta, \zeta, \phi, \phi' \;\rightarrow\; -Q, -f, -\mathbf{u}, -\delta, -\zeta, \phi, \phi' \;,$$

$$\mathbf{u}_1, \delta_1, \zeta_1, \phi_1' \;\rightarrow\; \mathbf{u}_1, \delta_1, \zeta_1, -\phi_1' \;,$$

$$\mathbf{u}_2, \delta_2, \zeta_2, \phi_2' \;\rightarrow\; -\mathbf{u}_2, -\delta_2, -\zeta_2, \phi_2' \;, \ldots$$

and so on, with alternating signs for even and odd subscripts. Under the sign changes, the diagnostic estimates for the various time derivatives behave in the same way as the actual time derivatives under PV-conserving balanced evolution. This reflects the fact that, after sign reversal, the new Q pattern is advected by the new $\mathbf{u}$ field in exactly the same way as the original Q pattern by the original $\mathbf{u}$ field, but with time running backwards, and that the same applies to the diagnostic estimates of the advection.

The inversion equations are nonlinear and implicit, and have to be solved iteratively. The numerical method is described in Appendix A.

4. A first example with mean depth 1 km

It will be recalled that the example of Fig. 1 has mean depth $\hat{h} = g^{-1}\hat{\phi} = 2\,\mathrm{km}$, and that its local Froude number, which in the present notation is

$$F = |\mathbf{u}|/(\hat{\phi} + \phi')^{1/2} \;, \tag{4.1}$$

reaches a maximum value $F_{\max} \simeq 0.5$ in the jet. This is hardly small in comparison with unity, and so the case of Fig. 1 is already a severe test of the PV inversion concept. But one would like to test the accuracy of the inversion operators in examples where the layer depth is still smaller and $F_{\max}$ still larger.

We now present some direct PV inversions in an example with mean depth 1 km. Fig. 2 shows the free-surface elevation $g^{-1}\phi'$ and velocity vectors $\mathbf{u}$ (Fig. 2a), divergence field δ (Fig. 2b), PV field Q (Fig. 2c), and local Froude number F (Fig. 2d). These fields are taken from a primitive equation integration like that of Fig. 1a–d, though at triangular truncation T63 rather than T106. The way in which the flow was set up is described in Appendix C. The minimum layer depth is less than 0.5 km, occurring at 90°W within the circumpolar vortex, counting 0°W at the bottom; see Fig. 2a. In Fig. 2d, values $F > 0.5$ are shaded. We see that $F_{\max} > 0.7$.

Fig. 3a shows the surface-elevation and velocity fields from a first order direct inversion of the PV distribution in Fig. 2c. The surface-elevation field differs from that in Fig. 2a by at most one contour. The second order direct inversion of the same PV distribution reconstructs the surface-elevation and velocity vectors much more accurately, producing a map (not shown) that is almost indistinguishable by eye from Fig. 2a. Fig. 3b presents the divergence field from the second order direct inversion; it differs from that in Fig. 2b by at most one contour. Fig. 3c presents the corresponding divergence difference field, defined as the divergence field of the primitive equation integration (Fig. 2b) minus the divergence field produced by the second order inversion (Fig. 3b), plotted using one tenth the contour interval.

Fig. 3d uses the same, finer, contour interval to show the difference between the divergence field of Fig. 2b and that of the third order direct inversion. Third order does not appear to be a dramatic improvement on second order, especially in the tropics; but we shall find in section 6 below that this is not due to the limitations of inversion, but rather to imbalance in the initial conditions for the primitive-equation integration. Indeed, in the second and third order divergence difference fields, Figs. 3c and 3d, one can see what appears to be the signature of a freely propagating tropical Kelvin wave. Closer analysis (see N88) confirms this interpretation. There are also, however, synoptic scale features in Figs. 3c and 3d, which quite unlike the Kelvin wave are phase-locked to troughs and ridges in the jet. Further discussion is deferred until section 6.

We have tried pushing $F_{\max}$ even higher by further reducing $\hat{h}$, with apparently successful results up to $F_{\max}$ values > 1.2 (N88, p. 98). But convergence was less sure; therefore we are not completely confident of numerical integrity of these results and will leave them aside here. We now turn to the normal mode approach.

5. Normal mode inversion

Nonlinear normal mode methods have been widely and successfully used in the initialization of global forecast models (e.g., Daley 1981, Errico 1982, & refs). It is next shown how the same methods can be used to construct another hierarchy of PV inversion operators, to be termed "normal mode" inversion operators. When judged in terms of cumulative accuracy (section 6), such inversion operators will prove to be the most accurate we have been able to construct.

The basic mathematical tool is the complete set of normal modes of the primitive equations linearized about rest, neglecting dissipation. For our hemispherical model domain these modes are the Hough functions with appropriate cross-equatorial symmetry or antisymmetry. They are classified into Rossby or slow modes on the one hand, and gravity or fast modes, including tropical Kelvin waves, on the other. If, in a numerical model of finite resolution, there are $2M$ degrees of freedom for each scalar field (e.g., $2M$ grid points, or M complex spectral coefficients), then there are $2M$ slow modes and $4M$ fast modes. Because of completeness, linear combinations of the normal modes can be used to represent the model fields. Hence the primitive equations (3.1) or (A.1) and domain boundary conditions are equivalent to a set of $3M$ ordinary differential equations of the form

$$\frac{d\mathbf{A}}{dt} + i\boldsymbol{\Lambda}\mathbf{A} = \mathbf{N}(\mathbf{A}, \mathbf{a}) \ , \qquad (5.1a)$$

$$\frac{d\mathbf{a}}{dt} + i\boldsymbol{\lambda}\mathbf{a} = \mathbf{n}(\mathbf{A}, \mathbf{a}) \ , \qquad (5.1b)$$

where $\mathbf{A}$ is a vector of length M consisting of the complex coefficients multiplying the slow modes, $\mathbf{a}$ is a vector of length $2M$ consisting of the complex coefficients multiplying the fast modes, $\boldsymbol{\Lambda}$ and $\boldsymbol{\lambda}$ are diagonal matrices of the (real) eigenfrequencies of the respective normal modes, and $\mathbf{N}$ and $\mathbf{n}$ are vectors consisting of the projections of the nonlinear terms onto the respective normal modes. $\mathbf{A}$ and $\mathbf{a}$ are functions of time only.

The first member of the new hierarchy, to be called "first order normal mode inversion," is defined as follows. Rather than deleting the time derivative in the divergence equation, as in the balance condition $(3.2a, b)$, we delete the time derivative in $(5.1b)$ to give

$$\mathbf{a} = (i\boldsymbol{\lambda})^{-1}\mathbf{n}(\mathbf{A}, \mathbf{a}) \ , \tag{5.2}$$

where the superscript -1 indicates the matrix inverse. This is the approximate balance condition imposed in the nonlinear normal-mode initialization scheme of Machenhauer (1977). It can be regarded as an implicit set of equations for $\mathbf{a}$ in terms of $\mathbf{A}$, expressing the idea that the imposition of balance means slaving the fast modes to the slow modes. The corresponding inversion operator is defined by (5.2) together with $(3.1d)$, or equivalently $(3.2c)$, after substitution of the relevant linear combinations of normal modes including that representing the prescribed PV distribution.

The second member of the hierarchy, to be called "second order normal mode inversion," is defined by equation $(3.1d)$ and the equations

$$\mathbf{A}_1 + i\boldsymbol{\Lambda}\mathbf{A} = \mathbf{N}(\mathbf{A}, \mathbf{a}) \ , \tag{5.3a}$$

$$\mathbf{a}_1 + i\boldsymbol{\lambda}\mathbf{a} = \mathbf{n}(\mathbf{A}, \mathbf{a}) \ , \tag{5.3b}$$

$$i\boldsymbol{\lambda}\mathbf{a}_1 = \mathbf{n}_1(\mathbf{A}, \mathbf{A}_1, \mathbf{a}, \mathbf{a}_1) \ . \tag{5.3c}$$

As in section 3, the notation $(\ )_n$ means a diagnostic estimate of the nth time derivative, to be solved for as part of the inversion operation, which is again, by definition, a purely diagnostic computation. The notation $\mathbf{n}_1(\mathbf{A}, \mathbf{A}_1, \mathbf{a}, \mathbf{a}_1)$ means the diagnostic estimate of the first time derivative of $\mathbf{n}(\mathbf{A}, \mathbf{a})$ obtained by differentiating $\mathbf{n}(\mathbf{A}, \mathbf{a})$ with respect to time and then substituting the diagnostic estimates $\mathbf{A}_1$ and $\mathbf{a}_1$ for the respective time derivatives $d\mathbf{A}/dt$ and $d\mathbf{a}/dt$. Eqs. (5.3) define the balance condition introduced by Tribbia (1984) to improve on the accuracy of Machenhauer's initialization scheme; (5.3) can again be regarded as an implicit set of equations for $\mathbf{a}$ (and the auxiliary variables $\mathbf{a}_1$, $\mathbf{A}_1$) in terms of $\mathbf{A}$. The corresponding inversion operator is defined by (5.3) together with $(3.1d)$.

The kth member of the hierarchy, to be called "kth order normal mode inversion," is defined for $k \geq 3$ by straightforwardly extending the same pattern (following Tribbia, op. cit.). The equations are

- Eq. (5.1a) and its time derivatives up to the $(k-2)$th, (5.4a)
- (5.1b) and its time derivatives up to the $(k-1)$th, with the term $d^k\mathbf{a}/dt^k$ deleted, (5.4b)
- Eq. (3.1d), after substitution of the relevant linear combinations of normal modes. (5.4c)

Eqs. (5.4a, b) define the balance condition and (5.4a–c) the inversion operator. It is understood that $d\mathbf{A}/dt$, $d^2\mathbf{A}/dt^2, ..., d^{k-1}\mathbf{A}/dt^{k-1}$ and $d\mathbf{a}/dt$, $d^2\mathbf{a}/dt^2, ..., d^{k-1}\mathbf{a}/dt^{k-1}$ are everywhere replaced by corresponding diagnostic estimates $\mathbf{A}_1$, $\mathbf{A}_2, ..., \mathbf{A}_{k-1}$ and $\mathbf{a}_1$, $\mathbf{a}_2, ..., \mathbf{a}_{k-1}$, to be treated as auxiliary variables and solved for as part of the inversion operation, as before. Note that the deletion of one time derivative is sufficient here, as compared with two in a direct inversion scheme, because the vector $\mathbf{a}$ has twice the length of the vector $\mathbf{A}$.

All these sets of inversion equations are nonlinear and implicit, as (3.1d), (5.2) and (5.3) illustrate, and iterative solution is again necessary. Further details of the method used can be found in Appendix B and in N88.

Note that the kth order normal mode inversion defined here is not closely equivalent to the kth order direct inversion defined in section 3, because the use of normal modes impresses a different structure on the inversion process, especially as regards latitudinal dependence on the hemisphere. At first order, for instance, the normal mode inversion takes some account of divergence, through the fast coefficients $\mathbf{a}(t)$, but direct inversion (3.2) does not, because of (3.2b). First order normal mode inversion more closely resembles a cut-down version of the second order direct inversion operator (3.3), in which (3.3d, e) are omitted along with the terms in $\mathbf{u}_1$ in (3.3b). The latter terms improve the performance in the tropics.

6. Cumulative accuracy of balanced model integrations

We now turn to the most stringent accuracy tests, namely, the results from 10-day integrations of PV-conserving balanced models in a case with mean depth $\hat{h} = g^{-1}\hat{\phi} = 1\,\mathrm{km}$, and $F_{\max} > 0.7$. This provides a severe, end-to-end test of cumulative accuracy over times that are substantial in comparison with eddy turnaround times. The test is all the more severe in that the case considered exhibits noticeable initial-condition sensitivity, or phase-space hyperbolicity.

The initial condition is the PV distribution shown in Fig. 6c. The definition of "PV-conserving balanced model" is that given in section 2. The prognostic equation is Eq. (2.1), i.e.,

$$\frac{\partial Q}{\partial t} = -\mathbf{u} \cdot \nabla Q \ , \tag{6.1}$$

plus a small ∇^6 hyperdiffusion term to control numerical noise at the truncation scale. T63 truncation is again used.

Since accuracy is to be judged by comparison with primitive equation evolution, we first present, in Fig. 4, a sequence of PV distributions from a ten-day integration of the primitive equation model in which the initial conditions are precisely those shown in Fig. 2a,b. Days are now re-numbered from zero. Note that, by day 2, the tongue of high PV that was situated at 100°E at day 0, i.e., pointing to the right in Fig. 2c, has become separated from the main vortex (by vortex rollup together with the action of the small model hyperdiffusion) to form a second cutoff vortex at 70°E. By day 6 the original cutoff vortex in the quadrant 0–90°W has re-merged with the main vortex. Meanwhile the second cutoff vortex has approached the equator and become somewhat elongated in shape. At day 8 this cutoff vortex is beginning to re-merge with the main vortex, a process that is complete by day 10.

Such vortex interaction processes are generally sensitive to initial conditions; and we shall see that in the present case the sensitivity is increased by the use of a hemispherical model domain. Fig. 5 shows the PV distribution at days 6 and 10 from another integration of the primitive equations with almost the same initial conditions as those for Fig. 4. The only difference is that the initial fields are those produced by a third order direct inversion of the PV distribution shown in Fig. 2c.

Several differences are immediately noticeable. They can largely be traced to the small, but important, advective influence of the tropical Kelvin wave that is present (but hardly visible, *cf.* Figs. 3c,d) in the fields shown in Fig. 2, but absent from the corresponding fields derived from the third order direct inversion. These differences in the initial conditions are enough to give rise to small, but important, differences in the positioning of the cutoff vortex near the equator around day 6 (compare Figs. 4c and 5a).

Figure 6 presents the PV fields at day 6, on the left of the figure, and at day 10, on the right, from integrations using the PV-conserving balanced models corresponding to first, second and third order direct inversion, proceeding downwards in the figure. The sensitivity connected with the positioning of the second cutoff vortex shows up immediately and more acutely. The first order balanced integration, shown in Fig. 6a, has this vortex at 5°W, 20°N. The second and third order balanced integrations, shown in Figs. 6c,e, have it much closer to the equator and successively further east. The second order balanced integration gives it a more elongated shape, as in day 6 of the first primitive equation integration (Fig. 4c). The varying positions and shapes of this cutoff vortex at day 6 lead to considerable differences in the subsequent re-merging with the main vortex. This is the main reason for the different PV distributions at day 10 shown in Figs. 6b,d,f. Comparison with Figs. 4 and 5 reveals the curious fact that it is the second order balanced integration that is closest to the first primitive equation integration (Fig. 4); however, the third order balanced integration does better by comparison with the second, re-initialized, primitive equation integration (Fig. 5), and this is the more meaningful comparison.

Fig. 7 is the same as Fig. 6 except that the balanced models are those associated with the first, second and third order normal mode inversion operators. The second and third order balanced integrations are in astonishingly close agreement with each other, and with the second, re-initialized, primitive equation integration (Fig. 5).

The essential findings, then, are (a) that the balanced-model behavior approaches the behavior of the re-initialized primitive equation integration (Fig. 5) much more closely than that of the first such integration shown in Fig. 4, and (b) that, when judged by comparison with the re-initialized primitive equation integration, the normal-mode inversions are the most accurate of those considered so far.

It is worth returning briefly to the origin of the sensitivity associated with the position of the cutoff vortex near the equator. Because of the hemispherical symmetry/antisymmetry (Appendix A), equivalent to the existence of a solid, though slippery, wall at the equator, there is an image vortex having opposite signed PV on the other side of the equator. If the vortex gets close enough to the equator, it will feel the induced fields from the image vortex sufficiently strongly that the pair of vortices can propagate eastward along the equator, as suggested by the wakes left behind the cutoff vortex at day 6 in Figs. 4c, 5a, 6c,e, and 7c,e. However, if the cutoff vortex does not reach this critical distance from the equator, it is advected westward by the easterly winds, the extreme example being Fig. 6a.

Figs. 8a and 8b summarize and quantify the differences between the model integrations. Fig. 8a shows graphs of the r.m.s. differences in the PV distributions, as a function of time, between the re-initialized primitive equation integration of Fig. 5 on the one hand, and the various balanced integrations of Figs. 6 and 7 and the first primitive equation integration of Fig. 4 on the other, the latter marked "PE". Fig. 8b shows the corresponding graphs for the r.m.s. surface-elevation differences. The two sets of graphs give similar impressions as to the relative accuracies of the different balanced models, as measured against re-initialized primitive equation behavior. They confirm that the least accurate balanced models are the first order direct inversion model followed by the first order normal mode model, and that the most accurate are the second and third order normal mode models. The last of these gives, at day 10, r.m.s. a surface-elevation difference just under 15 m, which may be compared with the persistence value of 206 m.

7. Galilean invariant PV inversion operators

Ideally, an operation like PV inversion ought to be Galilean invariant in the appropriate generalized sense, i.e., independent of any admissible motion of the coordinate frame. This is not true of any of the inversion operators constructed so far — by us, or, to our knowledge, by other investigators. One way to make an inversion operator Galilean invariant would be to formulate it entirely in terms of material or Lagrangian time derivatives $D/Dt = \partial/\partial t + \mathbf{u}\cdot\nabla$, as for instance in an initialization problem discussed by Hinkelmann (1969), rather than in terms of the contributions $\partial/\partial t$ and $\mathbf{u}\cdot\nabla$ separately. However, accuracy is lost when we neglect the $\mathbf{u}\cdot\nabla\mathbf{u}$ contribution to $D\mathbf{u}/Dt$, relegating gradient-wind-type corrections to higher orders. An alternative approach is to take an inversion operator like those already constructed, which are not intrinsically Galilean invariant but which might well be more computationally efficient, convenient, and accurate, and then to minimize some appropriate error norm over Galilean transformations compatible with the boundary conditions. For the hemisphere the admissible transformations are simply changes in the rotation rate Ω of the reference frame. Other Galilean transformations are inadmissible in this case, because of the inherently nonlocal nature of the inversion operation and the presence of the equatorial boundary of the hemispherical domain.

We test the feasibility of this idea by performing PV inversions in reference frames whose angular velocities Ω are both faster than, and slower than, the earth's angular velocity $\Omega = \Omega_{\mathrm{E}} = 7.3\times10^{-5}\mathrm{s}^{-1}$. In the comparison it is necessary that all the calculations refer to the same physical system with the same lower boundary, regardless of reference frame. When $\Omega = \Omega_{\mathrm{E}}$ we use Eqs. (3.1), (A.1), and the inversion equations as they stand. This implies, by the usual conventions in deriving (3.1) and (A.1), that the centrifugal potential appropriate to $\Omega = \Omega_{\mathrm{E}}$ has been absorbed into the gravitational potential (with the spheroidal-coordinate metric then approximated as spherical; see Phillips 1973). Thus the lower boundary is a level surface of the resulting total potential with $\Omega = \Omega_{\mathrm{E}}$. In order to describe the same physical system in a frame rotating with a different angular velocity $\Omega \neq \Omega_{\mathrm{E}}$, we can use the same equations with the different Ω value provided that we add to the right-hand side of the divergence equation $(A.1b)$ an extra term $(\Omega^2 - \Omega_{\mathrm{E}}^2)(3\sin^2\theta - 1)$, where θ is latitude, and make the corresponding modifications to the inversion equations. The extra term is just the divergence of the horizontal component of the change in the centrifugal force per unit mass, sometimes called the "non-Doppler effect". For direct inversion, only the first equation of each set is affected, e.g., $(A.3a)$, because $\partial/\partial t\{(\Omega^2 - \Omega_{\mathrm{E}}^2)(3\sin^2\theta - 1)\} = 0$.

Fig. 9a presents the divergence difference field from the third order direct inversion of the PV in Fig. 2c in a frame rotating with $\Omega = 7.8\times10^{-5}\mathrm{s}^{-1}$. In this reference frame, which is rotating slightly faster than the earth, by $0.5\times10^{-5}\mathrm{s}^{-1}$, the westerlies around the circumpolar vortex appear weaker and the tropical easterlies stronger. The increase in Ω corresponds to an equatorial velocity of $32\,\mathrm{ms}^{-1}$. Fig. 9b is for the opposite case of a frame rotating slightly slower by $0.5\times10^{-5}\mathrm{s}^{-1}$, i.e., rotating at $\Omega = 6.8\times10^{-5}\mathrm{s}^{-1}$, making the westerlies appear stronger and the easterlies weaker. One would expect the inversion errors from deleting partial time derivatives to increase where the winds are stronger and decrease where they are weaker; and this is just what is seen in Fig. 9a,b.

Fig. 10 summarizes the effect of changing the reference frame upon the various inversion operators, measured as in Fig. 4 by the r.m.s. synoptic-scale divergence error, counting all contributions whose total wavenumber $\geq$ 5. Results are plotted for the direct and normal mode inversions of the PV in Fig. 2c, computed in five reference frames whose angular velocities range from $\Omega = 6.3\times10^{-5}\mathrm{s}^{-1}$ to $\Omega = 8.3\times10^{-5}\mathrm{s}^{-1}$. We see that all the normal mode and higher order direct inversion operators have minimal error, by this criterion, in reference frames whose Ω values are close to the earth's value $\Omega = \Omega_{\mathrm{E}} = 7.3\times10^{-5}\mathrm{s}^{-1}$. The second and third order normal mode operators appear particularly sensitive to the changes in Ω, giving a sharp minimum.

The foregoing does not in itself provide a complete basis for constructing Galilean invariant inversion operators, because comparison with a primitive equation integration were used to estimate the errors. To be acceptable as a means of defining a PV inversion operation, which is a purely diagnostic operation, the process of minimizing the error over Ω calls for the use of an internal, likewise purely diagnostic, estimate of the error.

One way of obtaining such an estimate is by considering the behavior of the associated PV-conserving balanced model in the neighborhood of the diagnostic time. This can provide estimates of the time derivatives, such as $\partial^2\delta/\partial t^2$, $\partial^3\delta/\partial t^3$ or $d^2\mathbf{a}/dt^2$, $d^3\mathbf{a}/dt^3$, that were deleted in the process of constructing the inversion operator. Their smallness can be regarded as measuring the degree of self-consistency of balanced model evolution as an approximation to near-balanced primitive equation evolution, hence as measuring the accuracy of the inversion. Other choices are of course possible.

Fig. 11 shows how three such quantities vary with Ω. For illustrative purposes we take the case of third order direct inversion and its PV-conserving balanced model. The triangles are the r.m.s. values of an estimate of $\partial^2\delta/\partial t^2$ from the second time difference of δ, calculated by timestepping the balanced model forward and backward by one timestep. This is an internal error estimate in the sense required, making no reference to a primitive-equation integration. The circles are for a different such estimate, the same derivative $\partial^2\delta/\partial t^2$ calculated from the first centered time difference of δ_1. Both the absolute values, and the differences between the two, can be regarded as measures of inconsistency with primitive-equation evolution, hence inaccuracy. The squares are an estimate of $\partial^3\delta/\partial t^3$ from the second time difference of δ_1, multiplied by 10^3 s. All these internal estimates follow the trends in Fig. 10 for the third order direct inversions (solid squares). It seems clear that any of them could be used as the norm to be minimized in order to make the third order direct inversion operator into a Galilean invariant inversion operator.

Maps of the two diagnostic estimates of $\partial^2\delta/\partial t^2$ in the different rotating reference frames (not shown) display much the same spatial patterns and Ω dependence as the maps of synoptic-scale divergence differences exemplified by Fig. 9.

In our example it appears that the optimal Ω value is very close to that of the earth, Ω_{E}. It is clear from the foregoing, including Fig. 9, that this has no absolute significance: it comes about merely because the example happens to have regions both of easterlies and of westerlies. We may anticipate that in a hemispheric model of the wintertime upper stratosphere, for instance, the predominance of strong westerlies would dictate an optimal Ω value somewhat greater than Ω_{E}.

There are many more ways of constructing norms whose minimization over Ω should lead to Galilean invariant inversion operators. A thorough exploration of the possibilities is beyond the scope of this paper. For example, instead of evolving the balanced model, one could simply compare diagnostic estimates from successive orders of an inversion-operator hierarchy. Thus for instance one could take the r.m.s. difference between the δ_{k-2} values from the kth and $(k+1)$th direct inversion operators, or between the $\mathbf{a}_{k-1}$ values from the kth and $(k+1)$th normal mode inversion operators. These can be expected to give much the same kind of results as in Fig. 11. More sophisticated versions might try, for instance, to use norms designed to be sensitive to the disagreement between the diagnostic estimates of $\mathbf{u}\cdot\nabla Q$ given by two successive orders of inversion operator. These should give a measure of accuracy having direct relevance to the accuracy with which the corresponding PV-conserving balanced model advects the PV, hence relevance to the cumulative accuracy of that balanced model.

8. Concluding remarks

The astonishing cumulative accuracies at $F \lesssim 0.7$ implied by Figs. 7 and 8 may have distracted attention from the question of local mass conservation mentioned in section 1. Primitive-equation evolution conserves mass locally, meaning that Eq. (3.1c) is satisfied exactly, by definition, apart from numerical discretization errors. The question is whether the velocity and surface elevation fields produced by our PV-conserving balanced models likewise satisfy Eq. (3.1c) exactly, rather than merely to excellent approximation as the foregoing results suggest. This is an important question of principle. Although the PV-conserving balanced models are by construction globally mass-conserving — constancy of the mean depth $\hat{h} = g^{-1}\hat{\phi}$ is imposed computationally — there is no reason to suppose that they conserve mass locally, nor, for that matter, energy and momentum either locally or globally. Recent work at the Newton Institute in collaboration with Drs E. Neven, S. Ren, and I. Roulstone, as yet unpublished, has shown that the balanced models associated with our direct inversion operators do indeed violate local mass conservation; and the same must, presumably, hold for our normal mode inversion operators. It turns out to be simple to modify the second-order direct inversion operator to conserve mass locally,[5]

[5]First substitute (3.3c) into (3.3e) then delete $\mathrm{div}^{-1}\delta$ from (3.3c). This makes (3.3) with $DQ/Dt = 0$ into the local-mass-conserving version of the Bolin–Charney balance equations, as discussed e.g. by Gent and McWilliams (1984) and Whitaker (1993).

but impossible, as far as we can see, to achieve any such modification at higher order. Yet the cumulative accuracy evidenced most strikingly in the case of Fig. 7f not only provides a powerful check on the concepts and numerical codes used in this study, but also implies that departures from local mass conservation must be tiny in such cases, even though nonzero.

Whether local mass conservation is *desirable* as a property of balanced models is another question again. It is strongly arguable that enforcement of exact local mass conservation, and indeed energy and momentum conservation, would be likely to degrade the accuracy of a PV-conserving balanced model. In primitive-equation evolution, the spontaneous-adjustment emission of inertia–gravity waves — involving the spontaneous mutual adjustment of the mass and velocity fields within an unsteady, freely-evolving vortical flow (Ford et al. 1999 & refs.) — must modify the local mass, energy, and momentum budgets in ways that cannot be perfectly captured by a balanced model. Mass adjustments or rearrangements on the timescales of fast gravity-wave motion, in primitive-equation evolution, might be partially mimicked in an accurate balanced model as instantaneous mass rearrangements. By definition, such rearrangements require infinite velocities, and so cannot be exactly compatible with local mass conservation described by a velocity field that remains finite. Something has to give way.

In this connection, recent results on Hamiltonian balanced models, discussed and further developed at the Newton Institute Programme, seem to us to be very interesting. There are three essential points. First, it is now known how, in principle, to construct the Hamiltonian balanced model corresponding to any given balance condition, no matter how accurate — for instance the balance condition defined by equations $(3.6a)$–$(3.6i)$ above, or by $(5.3a)$–$(5.3c)$ or by $(A.3a)$–$(A.3i)$. That knowledge has emerged from the work of Salmon (1983, 1985, 1988), Allen and Holm (1996), Roulstone and Sewell (1996), and McIntyre and Roulstone (1996). Second, any balanced model so constructed will automatically respect mass, energy and momentum conservation, as well as PV conservation, in virtue of its Hamiltonian structure.

Third, however, such exact conservation comes at a price: the model's velocity field is always 'split' into two separate velocity fields, one of which advects materially conserved quantities like PV as well as the fluid particles themselves (thus satisfying local mass conservation exactly), and the other of which is the velocity field from which the PV, and the energy and momentum, are evaluated. The latter quantities, when so evaluated, have the exactly the same formulae as for the primitive equations that the balanced model seeks to approximate. This explains, incidentally, the peculiar formula for the semigeostrophic PV discovered by Hoskins (1975). Thus, e.g., for shallow water systems, the PV is Rossby's exact PV (Rossby 1936), formula $(3.1d)$ above, provided that it is evaluated from the second, not the first, of the model's two velocity fields.

It is still true, then, that 'something has to give way'. Whenever one insists on full and exact conservation properties by imposing Hamiltonian structure, what gives way is the uniqueness of the velocity field. A corollary is that, because PV-conserving balanced models of the type studied here are defined to have unique velocity fields, they cannot be Hamiltonian.

Acknowledgments.

We thank J. S. Allen, R. Bleck, O. Bühler, S. P. Cooper, M. J. P. Cullen, C. A. Davis, D. G. Dritschel, K. A. Emanuel, R. Ford, P. A. Glendinning, P. H. Haynes, B. J. Hoskins, I. N. James, M. N. Juckes, A. R. Mohebalhojeh, E. C. Neven, S. Ren, I. Roulstone, T. G. Shepherd, A. J. Simmons, G. J. Shutts, and J. J. Tribbia for helpful comments and correspondence. Two anonymous reviewers provided further perceptive and helpful comments. This work received support from the Commonwealth Scholarship Commission, the Cambridge Commonwealth Trust, the US Office of Naval Research, and the UK Natural Environment Research Council through the UK Universities' Global Atmospheric Modelling Project (grant GR3/6516) and through the British Antarctic Survey (grant GST/02/446), and more recently from the Isaac Newton Trust, the Isaac Newton Institute for Mathematical Sciences, and a SERC/EPSRC Senior Research Fellowship to MEM. We are especially grateful to B. J. Hoskins, I. N. James and M. N. Juckes for generous help with the numerical models used for the primitive equation integrations, which were based on those developed in the Meteorology Department at the University of Reading. S. P. Cooper and C. J. Cooper gave expert help with the task of making high-resolution model output intelligible to the human eye, and in many other ways.

APPENDIX A

Direct inversion on a hemisphere

In this appendix we define the direct inversion operators on the hemisphere, and describe the numerical procedure used to solve the resulting equations. The primitive equations for the shallow water system on a hemisphere are

$$\frac{\partial \zeta}{\partial t} + 2\Omega\mu\delta + \frac{2\Omega V}{a} \;=\; N_\zeta, \tag{A.1a}$$

$$\frac{\partial \delta}{\partial t} + \nabla^2\phi' - 2\Omega\mu\zeta + \frac{2\Omega U}{a} \;=\; N_\delta, \tag{A.1b}$$

$$\frac{\partial \phi'}{\partial t} + \hat{\phi}\delta \;=\; N_\phi, \tag{A.1c}$$

where Ω is the Earth's angular velocity, a is the Earth's radius, $\mu = \sin\theta$ where θ is latitude,

$$(U, V) \;=\; \mathbf{u}\cos\theta \;=\; (1-\mu^2)^{1/2}(\mathbf{k}\times\nabla\psi \;+\; \nabla\chi)\ ,$$

i.e.,

$$U \;=\; \frac{1}{a}\left(\frac{\partial\chi}{\partial\lambda} - (1-\mu^2)\frac{\partial\psi}{\partial\mu}\right)\ , \qquad V \;=\; \frac{1}{a}\left(\frac{\partial\psi}{\partial\lambda} + (1-\mu^2)\frac{\partial\chi}{\partial\mu}\right)$$

where λ is longitude. The nonlinear terms on the right of Eqs. $(A.1)$ are

$$N_\zeta \;=\; -\frac{1}{a}\left(\frac{1}{1-\mu^2}\frac{\partial}{\partial\lambda}(U\zeta) + \frac{\partial}{\partial\mu}(V\zeta)\right), \tag{A.2a}$$

$$N_\delta \;=\; \frac{1}{a}\left(\frac{1}{1-\mu^2}\frac{\partial}{\partial\lambda}(V\zeta) - \frac{\partial}{\partial\mu}(U\zeta)\right) - \tfrac{1}{2}\nabla^2\left(\frac{U^2+V^2}{1-\mu^2}\right), \tag{A.2b}$$

$$N_\phi \;=\; -\frac{1}{a}\left(\frac{1}{1-\mu^2}\frac{\partial}{\partial\lambda}\left(U\phi'\right) + \frac{\partial}{\partial\mu}\left(V\phi'\right)\right), \tag{A.2c}$$

and all other quantities are as defined in section 3; in particular,

$$\nabla^2\psi \;=\; \zeta \qquad \text{and} \qquad \nabla^2\chi \;=\; \delta\ .$$

The expression $(A.2b)$ may be compared to the right-hand side of $(3.1b)$ with $-\nabla.(\mathbf{u}.\nabla\mathbf{u})$ rewritten as $-\nabla.(\zeta\,\mathbf{k}\times\mathbf{u}) - \frac{1}{2}\nabla^2(|\mathbf{u}|^2)$.

The direct inversion operators are constructed as in section 3. For a pth order operator, we take Eqs. $(A.1)$ supplemented, when $p \geq 2$, by the $(p-1)$ equations $\partial(A.1b)/\partial t$, ..., $\partial^{p-1}(A.1b)/\partial t^{p-1}$, then delete the two time derivatives $\partial^{p-1}\delta/\partial t^{p-1}$ and $\partial^p\delta/\partial t^p$ and replace all the remaining time derivatives by diagnostic estimates. The latter are denoted as before by subscripts $1, 2, \ldots$. These equations define the balance condition; to get the PV inversion operator we append the formula defining the PV.

For instance the equations defining the third order direct inversion operator on a hemisphere, in which $\partial^2\delta/\partial t^2$ and $\partial\delta^3/\partial t^3$ are deleted, may be compared to Eqs. (3.6) on the f-plane and are

$$\nabla^2\phi' - 2\Omega\mu\nabla^2\psi + \frac{2\Omega U}{a} \;=\; N_\delta - \delta_1\ , \tag{A.3a}$$

$$\tilde{\mathcal{L}}\chi - \frac{2\Omega}{a^2}\frac{\partial\chi_1}{\partial\lambda} - \frac{4\Omega^2\mu}{a^2}\frac{\partial\psi}{\partial\lambda} \;=\; S\ , \tag{A.3b}$$

$$\tilde{\mathcal{L}}\chi_1 \;=\; T\ , \tag{A.3c}$$

$$(U,V) \;=\; (1-\mu^2)^{1/2}\,(\mathbf{k}\times\nabla\psi + \nabla\chi)\ , \tag{A.3d}$$

$$(U_1,V_1) \;=\; (1-\mu^2)^{1/2}\,(\mathbf{k}\times\nabla\psi_1 + \nabla\chi_1)\ , \tag{A.3e}$$

$$(U_2,V_2) \;=\; (1-\mu^2)^{1/2}\,\mathbf{k}\times\nabla\psi_2\ , \tag{A.3f}$$

$$\zeta_1 + 2\Omega\mu\,\delta + \frac{2\Omega V}{a} \;=\; N_\zeta\ , \tag{A.3g}$$

$$\zeta_2 + 2\Omega\mu\,\delta_1 + \frac{2\Omega V_1}{a} \;=\; (N_\zeta)_1\ , \tag{A.3h}$$

$$\phi_1' + \hat{\phi}\,\delta \;=\; N_\phi\ , \tag{A.3i}$$

$$\nabla^2\psi - \frac{2\Omega\mu}{\hat{\phi}}\,\phi' \;=\; \frac{Q'}{g}(\hat{\phi}+\phi')\ , \tag{A.3j}$$

where $\tilde{\mathcal{L}}$ is the fourth-order linear operator

$$\tilde{\mathcal{L}} \;=\; \hat{\phi}\nabla^4 - 4\Omega^2\mu^2\nabla^2 - \frac{4\Omega^2\mu}{a^2}(1-\mu^2)\frac{\partial}{\partial\mu} \tag{A.4}$$

(fourth order because it acts on χ and χ_1 rather than on δ and δ_1), and where

$$S \;=\; \nabla^2 N_\phi - (N_\delta)_1 - 2\Omega\mu N_\zeta - \frac{2\Omega}{a^2}(1-\mu^2)\frac{\partial\psi_1}{\partial\mu}, \tag{A.5a}$$

$$T \;=\; \nabla^2(N_\phi)_1 - (N_\delta)_2 - 2\Omega\mu(N_\zeta)_1 - \frac{2\Omega}{a^2}(1-\mu^2)\frac{\partial\psi_2}{\partial\mu} + \frac{4\Omega^2\mu}{a^2}\frac{\partial\psi_1}{\partial\lambda}\ . \tag{A.5b}$$

The differential operators on the left-hand sides of Eqs. ($A.3a, b, c, j$) are all elliptic, and very well behaved numerically. Given the right-hand sides, therefore, we can solve robustly for (ϕ', χ, χ_1, ψ) under the conditions that ϕ', χ, and χ_1 are symmetric about the equator (as also are δ, δ_1, U, U_1, and U_2) and that ψ is antisymmetric (as also are ψ_1, ψ_2, ζ_1, ζ_2, V, V_1, V_2, and Q'). Global mass conservation requires that the area integral of ϕ' over the hemisphere is zero.

The whole system is solved by iteration, with the solution of the elliptic equations $(A.3a, b, c, j)$ as an inner loop. Outside this, we need to evaluate the diagnostic estimates $(N_\delta)_1$, $(N_\phi)_1$, $(N_\zeta)_1$, and $(N_\delta)_2$ at each stage. These nonlinear terms could have been written out explicitly as in Eqs. (3.6), and the corresponding code written from scratch. But in practice it is simpler and safer, and equivalent within an easily-controllable truncation error, to take advantage of the existing model code and evaluate these terms within the outer iteration loop by a procedure that we call "quasi-timestepping". This is to emphasize that it is still part of solving the inversion problem $(A.3)$, which by definition is a purely diagnostic process.

At each stage in the outer loop, the exact primitive equations $(A.1)$ are timestepped forward and backward, just once in each direction, using a second order Runge–Kutta scheme (local truncation error $O(\Delta t^3)$) with the current best-guess fields as initial conditions. This gives estimates for all the fields at, say, times t and $t \pm \Delta t$, with error $O(\Delta t^3)$. From these estimates, N_δ, N_ϕ, and N_ζ are evaluated at $t \pm \Delta t$, allowing $(N_\delta)_1$, $(N_\phi)_1$, $(N_\zeta)_1$, and $(N_\delta)_2$ to be estimated as centered time differences, the last with error $O(\Delta t)$. With sufficiently small Δt, it is straightforward to show by Taylor expansion that, provided the whole iterative process converges, it is equivalent, with error $O(\Delta t)$, to solving the foregoing equations with $(N_\delta)_1$, $(N_\phi)_1$, $(N_\zeta)_1$, and $(N_\delta)_2$ written out explicitly as in Eqs. (3.6). In particular, not only $(A.3d, e)$ and $(A.3g$–$i)$ but also $(A.3f)$ are satisfied, with error $O(\Delta t)$, as the following consideration shows. Denote by $\tilde{\delta}_2$ the second time difference of δ produced by the quasi-timestepping procedure just described. Even though $\tilde{\delta}_2$ will be nonzero during the interation process — with the effect of adding a term $\propto \nabla\tilde{\chi}_2 = \nabla(\nabla^{-2}\tilde{\delta}_2)$ to the right-hand side of $(A.3f)$ — convergence of the iteration will force $\tilde{\delta}_2$ to zero by virtue of $(A.3b)$, $(A.3g)$, and $(A.3i)$, which must all be satisfied, with error $O(\Delta t)$ at worst, once the iteration has converged. For with

$(A.3g)$ and $(A.3i)$ satisfied, $(A.3b)$ corresponds, by construction, to $-\partial(A.1b)/\partial t$ with $\partial^2\delta/\partial t^2$ replaced by zero.

In the numerical solution procedure, each field is expressed as the sum of spectral coefficients multiplied by spherical harmonic functions. By using standard recurrence relations, the left-hand sides of Eqs. $(A.3a, b, c, j)$ are written in the form of a (robustly nonsingular) constant-element matrix times a vector of spectral coefficients. This system of equations is block-diagonal, and can be partitioned according to values of the zonal wavenumber. A sparse matrix solver is used to solve it with the current best guess for the right-hand sides of $(A.3a, b, c, j)$. The new fields are then used to update the nonlinear terms and diagnostic time-derivative estimates. The iteration is continued until convergence is reached, the criterion for convergence being that the r.m.s. value of $\phi'_{(q)} - \phi'_{(q-1)}$ should be at most 10^{-7} of the r.m.s. value of $\phi'_{(q)}$, the qth iterate of the surface geopotential fluctuation. Typically, some 30 to 50 iterations were found to be necessary in a third order direct inversion. Improved convergence was obtained by first performing a few iterations of the first order direct inversion equations. For the small mean depths discussed here it was found necessary, in addition, to use under-relaxation of the nonlinear terms in order to obtain convergence. An under-relaxation parameter of 0.6 was used.

The coding error mentioned in the introduction produced incorrect values of one (fortunately small) spectral coefficient when, and only when, it entered the quasi-timestepping procedure for estimating $(N_\delta)_1$, $(N_\phi)_1$, $(N_\zeta)_1$, and $(N_\delta)_2$. The spectral coefficient affected was that of the $P_1(\mu)$ contribution to ψ, i.e. the first zonally symmetric and meridionally antisymmetric contribution to ψ, corresponding to solid rotation.

APPENDIX B

Normal mode inversion on a hemisphere

In this appendix we describe the numerical procedure used to carry out the kth order normal mode PV inversion defined by (5.4). First, the left hand side of the PV equation in the form (3.2c) is split into the slow and fast contributions from the normal mode superposition, giving, symbolically,

$$\left(\zeta - \frac{f\phi'}{\hat{\phi}}\right)_{\mathrm{slow}} = \frac{Q'}{g}(\hat{\phi} + \phi') - \left(\zeta - \frac{f\phi'}{\hat{\phi}}\right)_{\mathrm{fast}} . \qquad (B.1)$$

This is solved to find the slow-mode coefficients, given a previous guess for the slow-mode and fast-mode coefficients. Solution for the slow-mode coefficients, $\mathbf{A}$, is done in spherical harmonic spectral space, by substituting the slow-mode coefficients multiplied by the corresponding slow-mode eigenfunctions (expressed in terms of spherical harmonics) into the left hand side of Eq. (B.1), and equating this to a spherical harmonic spectral representation of the right hand side evaluated from the previous guess. This reduces the problem to a system of M equations for the M new slow-mode coefficients (where there are $2M$ degrees of freedom for each scalar field). This system of equations partitions with zonal wavenumber and can be expressed as a series of constant element, non-sparse matrices multiplied by the corresponding vector of slow-mode coefficients. The inverses of these non-sparse matrices need only be calculated once.

The wavenumber zero part of (B.1) is solved differently taking advantage of the fact that the slow-mode eigenfunctions with wavenumber zero are degenerate (all having eigenfrequency zero). These eigenfunctions all have nondivergent wind fields in geostrophic balance with their geopotential fields. Hence the wavenumber zero part of (B.1) is more simply solved by substituting the condition for geostrophic balance into the spectral form of (B.1); the new guess for the wavenumber zero slow-mode coefficients can then be determined by a sparse matrix inversion.

Next, the new slow-mode coefficients $\mathbf{A}$ are substituted into the ϕ' contribution on the right hand side of Eq. (B.1), keeping the old fast coefficients $\mathbf{a}$, and the above procedure iterated. This consists of an inner loop within the whole iteration involving (5.4). At each inner-loop iteration the right hand side of (B.1) is calculated with the latest $\mathbf{A}$

but fixed $\mathbf{a}$. Four to ten iterations are used. In inversions at small mean depths such as 1 km, where ϕ' can be a large fraction of $\hat{\phi}$, it has been found necessary to use under-relaxation of the right hand side.

With the resulting estimate for the slow-mode coefficients, $\mathbf{A}$, the algorithm then proceeds to update the fast-mode coefficients, $\mathbf{a}$, by using the balance condition. For example in Eq. (5.2), which is the case of $(5.4a, b)$ representing first order normal mode balance, the nonlinear terms N_ζ, N_δ and N_ϕ defined by $(A.2)$ are computed and their projection $\mathbf{n}$ onto the fast modes is evaluated from

$$n_i = \int_0^1 \int_0^{2\pi} \left(\Phi_i^* N_\phi - \hat{\phi}(\Psi_i^* \nabla^2 N_\zeta + \mathrm{X}_i^* \nabla^2 N_\delta) \right) d\lambda d\mu, \qquad (B.2)$$

where n_i is the ith component of $\mathbf{n}$, and $\Phi_i, \Psi_i, \mathrm{X}_i$ are the ϕ, ψ, χ components of the ith fast-mode eigenfunction normalized such that

$$\int_0^1 \int_0^{2\pi} \left(|\Phi_i^*|^2 + \hat{\phi}(|\nabla \Psi_i|^2 + |\nabla \mathrm{X}_i|^2) \right) d\lambda d\mu = 1;$$

the asterisks denote the complex conjugate. Equation $(B.2)$ is evaluated in spherical harmonic spectral space, which reduces it to evaluating a summation over total wavenumber. Since the right hand sides of (5.2) and its higher-order counterparts depend on the fast-mode coefficients, it is necessary to iterate here as well. But first the PV equation $(B.1)$ is solved again, repeating the inner loop iteration using the new fast coefficients $\mathbf{a}$. With the resulting updates of $\mathbf{a}$ and $\mathbf{A}$, new values for the right hand side of (5.2) are then calculated. This gives an improved estimate of $\mathbf{a}$; the whole iteration is then repeated.

For higher order balances the numerical procedure is the same as before, except that it is necessary to evaluate diagnostic estimates for the time derivatives of the nonlinear terms, as for instance on the right-hand side of (5.3). These are evaluated by quasi-timestepping in the sense explained in Appendix A. For third order normal mode inversion, in examples with small mean depth $g^{-1}\hat{\phi}$ such as 1 km, best convergence was achieved by ramping up through first then second order. The 1 km examples of section 6 used 8 first order, 8 second order, then 25 third order iterations. Under-relaxation on the fast-mode coefficients was also found to improve convergence.

APPENDIX C

Topographic forcing

The fields shown in Fig. 2 are for day 25 of a primitive-equation integration in which the motion was excited, prior to day 25, from zonally symmetric initial conditions, by a smoothly-varying artificial topographic forcing. The time-dependent bottom topography H is related to the geopotential ϕ and the layer depth h by $\phi = g(h + H)$. Therefore the only change needed in the primitive equations $(A.1)$ is to replace ϕ' by $\phi' - gH$ in the mass conservation equation $(A.1c)$. In the integrations shown, H depends on t, θ, λ according to

$$H = H_0 A(t) B(\theta) C(\lambda) \ , \tag{C.1}$$

where

$$A(t) = \begin{cases} \frac{1}{2}(1 - \cos\dfrac{t\pi}{T}) & t \leq 2T \text{ days} \ , \\ \\ 0 & t \geq 2T\text{days} \ , \end{cases} \tag{C.2}$$

$$B(\theta) = \frac{\cot^2\theta}{\cot^2\theta_0} \exp\left(1 - \frac{\cot^2\theta}{\cot^2\theta_0}\right) , \tag{C.3}$$

and

$$C(\lambda) = -(\cos\lambda + c_m \cos m\lambda) \ . \tag{C.4}$$

The origin of longitude, $\lambda = 0$, is plotted at "six o'clock" in the figures. The growth or decay time for the topography $T = 12$ days. In the 1 km mean depth integration of Fig. 2, $H_0 = 0.45$km, $\theta_0 = 35°$, $c_m = 0.2$, and $m = 5$. Note that the smooth function $B(\theta)$ peaks at $\theta = \theta_0$ with value $B(\theta_0) = 1$.

REFERENCES

Allen, J. S., 1993: Iterated geostrophic intermediate models. *J. Phys. Oceanog.*, **23**, 2447–2461.

——, and D. D. Holm, 1996: Extended-geostrophic Hamiltonian models for rotating shallow water motion. *Physica D*, **98**, 229-248.

Bolin, B., 1955: Numerical forecasting with the barotropic model. *Tellus*, **7**, 27–49.

Bühler, O., 1998: A shallow-water model that prevents nonlinear steepening of gravity waves. *J. Atmos. Sci.*, **55**, 2884–2891.

——, and M. E. McIntyre, 1998: On non-dissipative wave–mean interactions in the atmosphere or oceans. *J. Fluid Mech.*, **354**, 301–343.

Charney, J. G., 1948: On the scale of atmospheric motions. *Geofysiske Publ.*, **17(2)**, 3–17.

[[**Editorial office NB:** Please do not delete the "(2)", which is essential for unambiguous referencing in *Geofysiske Publikasjoner*.]]

——, 1955: The use of the primitive equations of motion in numerical prediction. *Tellus*, **7**, 22–26.

——, 1962: Integration of the primitive and balance equations. In *Proc. Int. Symp. Num. Wea. Pred.*, Tokyo, Meteorol. Soc. Japan, 131–152.

——, 1963: A note on large-scale motions in the tropics. *J. Atmos. Sci.*, **20**, 607–609.

Daley, R., 1981: Normal mode initialization. *Rev. Geophys. Space Physics*, **19**, 450–468.

Errico, R. M., 1982: Normal mode initialization and the generation of gravity waves by quasi-geostrophic forcing. *J. Atmos. Sci.*, **39**, 573–586.

Ford, R., M. E. McIntyre, and W. A. Norton, 1999: Balance and the slow quasi-manifold: some explicit results. J. Atmos. Sci., **56**, this issue. [[**EDITORS: We understand that this companion paper is to appear in the same issue of JAS; please insert page numbers if possible.**]]

Gent, P. R., and J. C. McWilliams, 1984: Balanced models in isentropic coordinates and the shallow water equations. *Tellus*, **36A**, 166–171.

Grønås, S., and M. A. Shapiro, 1999: *The Life Cycles of Extratropical Cyclones*, Boston, American Meteorological Soc., in press for March 1999.

[[**Editorial office: please check whether this volume (final form of the Bergen Symposium collection) has appeared and can be fully referenced.**]]

Haltiner, G. J., and R. T. Williams, 1980: *Numerical prediction and dynamic meteorology.* Wiley, New York, 477pp.

Haynes, P. H., and M. E. McIntyre, 1990: On the conservation and impermeability theorems for potential vorticity. *J. Atmos. Sci.*, **47**, 2021–2031.

Hinkelmann, K. H., 1969: Primitive equations. In: *Lectures in Numerical Short-range Weather Prediction. Regional Training Seminar, Moscow.* WMO No. **297**, pp 306–375.

Hoskins, B. J., 1975: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, **32**, 233–242.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential-vorticity maps. *Q. J. Roy. Meteorol. Soc.*, **111**, 877–946. Also **113**, 402–404.

Lynch, P., 1989: The slow equations. *Q. J. Roy. Meteorol. Soc.*, **115**, 201–219.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water equation model with applications to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253–271.

McIntyre, M. E., and W. A. Norton, 1990a: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. J. Fluid Mech, **212**, 403–435. (G. K. Batchelor Festschrift Issue.) *Corrigendum* **220**, 693.

——, and ——, 1990b: Nonlinear vorticity or potential vorticity inversion In: *Topological Fluid Mechanics*, ed. H. K. Moffatt and A. Tsinober; Cambridge University Press, 355–358.

——, and I. Roulstone, 1996: Hamiltonian balanced models: constraints, slow manifolds and velocity splitting. Forecasting Research Division, Scientific Paper No. **41**, 1–49, available from UK Meteorological Office, Bracknell, Berks. RG12 2SZ, UK, also from http://www.atmos-dynamics.damtp.cam.ac.uk/people/mem/

McWilliams, J. C., 1985: A uniformly valid model spanning the regimes of geostrophic and isentropic, stratified turbulence: balanced turbulence. *J. Atmos. Sci.*, **42**, 1773–1774.

Mo, R., O. Bühler, and M. E. McIntyre, 1998: Permeability of the stratospheric vortex edge: On the mean mass flux due to thermally dissipating, steady, non-breaking Rossby waves. *Q. J. Roy. Meteorol. Soc.*, **124**, 2129–2148.

Norton, W. A., 1988: Balance and potential vorticity inversion in atmospheric dynamics. University of Cambridge, PhD Thesis, 167 pp. ("N88" in text.)

Phillips, N. A., 1973: Principles of large-scale numerical weather prediction. In: *Dynamic Meteorology*, ed. P. Morel; Dordrecht, Reidel, 3–96.

Rossby, C. G., 1936: Dynamics of steady ocean currents in the light of experimental fluid mechanics. *Mass. Inst. of Technology and Woods Hole Oc. Instn. Papers in Physical Oceanography and Meteorology*, **5**(1), 1–43.

[[**Editorial office NB:** Please do not delete the "(1)", which is essential for unambiguous referencing in *Papers in Physical Oceanography and Meteorology*.]]

Roulstone, I., and M. J. Sewell, 1996: Potential vorticities in semi-geostrophic theory. *Q. J. Roy. Meteorol. Soc.*, **122**, 983–992.

Salmon, R., 1983: Practical use of Hamilton's principle. *J. Fluid Mech.*, **132**, 431–444.

——, 1985: New equations for nearly geostrophic flow. *J. Fluid Mech.*, **153**, 461–477.

——, 1988: Semigeostrophic theory as a Dirac-bracket projection. *J. Fluid Mech.*, **196**, 345–358.

Temperton, C., 1988: Implicit normal mode initialization. *Mon. Wea. Rev.*, **116**, 1013–1031.

——, 1989: Implicit normal mode initialization for spectral models. *Mon. Wea. Rev.*, **117**, 436–451.

Tribbia, J. J., 1984: A simple scheme for high-order nonlinear normal mode initialization. *Mon. Wea. Rev.*, **112**, 278–284.

Vallis, G. K., 1996: Potential vorticity inversion and balanced equations of motion for rotating and stratified flows. *Q. J. Roy. Meteorol. Soc.*, **122**, 291–322.

Warn, T., 1997: Nonlinear balance and quasigeostrophic sets. *Atmos.–Ocean*, **35**, 135–145. *Note: this paper was originally written in 1983.*

Warn, T., O. Bokhove, T. G. Shepherd, and G. K. Vallis, 1995: Rossby number expansions, slaving principles, and balance dynamics. *Q. J. Roy. Meteorol. Soc.*, **121**, 723–739.

Whitaker, J. S., 1993: A comparison of primitive and balance equation simulations of baroclinic waves. *J. Atmos. Sci.*, **50**, 1519–1530.

FIGURE CAPTIONS

FIG. 1. Diagnostic test of potential vorticity (PV) inversion, using a PV field produced by a primitive-equation shallow-water integration on a hemisphere. Panels (a)–(d) show the fields produced by the primitive-equation integration and (e),(f) a reconstruction of (a),(b) by inversion from the PV alone. The mean depth of the layer is 2 km; maximum jet velocity $\simeq$ 70 ms^{-1}; local Froude numbers, defined as $F = |\mathbf{u}|/(\hat{\phi} + \phi')^{1/2}$, attain maximum values $\simeq$ 0.5. Model resolution is spectral triangular truncation T106, i.e., the numerical models retains total spherical harmonic wavenumbers up to 106. Solid contours show positive values, long-dashed contours negative values, and dotted contours zero. The projection is polar stereographic. In detail: (a) Arrows show the velocity, $\mathbf{u}$, on the scale indicated; contours show free-surface elevation, i.e., g^{-1} times the departure ϕ' of the geopotential from its area mean value $\hat{\phi}$. The contour interval is 100 m. (b) Divergence, δ, contoured at intervals of 6×10^{-7} s^{-1}. (c),(d) PV, Q, defined as in Eq. (3.1d), absolute vorticity divided by layer depth: contours in (c) are at intervals of $1{\times}10^{-8}$ m^{-1} s^{-1}. The shading in the contour plot highlights Q values lying between 4 and 6 of these units. The grayscale representation (d) of the same information is monotonic from light to dark, increasing from zero at the equator to a maximum value of $1{\times}10^{-7}$ m^{-1} s^{-1} near the pole. (e),(f) As in (a),(b) but reconstructed from the PV alone using the third order direct inversion algorithm described in section 3 and Appendix A.

FIG. 2. Fields from a primitive-equation shallow water model integration on a hemisphere with mean depth of 1 km. Model resolution is spectral triangular T63 truncation, here and in subsequent figures. (a) Arrows show the velocity, $\mathbf{u}$, on the scale indicated; contours show free-surface elevation $g^{-1}\phi'$. The contour interval is 100 m. (b) Divergence, δ, contoured at intervals of 1×10^{-6} s^{-1}. (c) PV, Q, contoured at intervals of $3{\times}10^{-8}$ m^{-1} s^{-1}; the shading in the contour plot highlights values lying between 4 and 6 of these units. (d) Local Froude number defined as $F = |\mathbf{u}|/(\hat{\phi} + \phi')^{1/2}$, contoured at intervals of 0.1; the shading highlights values $F > 0.5$, and $F_{\max}$ is just over 0.7.

FIG. 3. (a) Contours of free-surface elevation, $g^{-1}\phi'$, and velocity, $\mathbf{u}$, reconstructed by first order direct inversion of the PV field shown in Fig. 2c. Contour interval is 100 m, as in Fig. 2a. (b) Divergence, δ, reconstructed by second order direct inversion of the PV field shown in Fig. 2c. Contour interval is 1×10^{-6} s^{-1}, as in Fig. 2b. (c) Divergence difference field defined as δ from the primitive-equation model (Fig. 2b) minus δ from the second order direct inversion (Fig. 3b). The contour interval is $1 \times 10^{-7}\,\mathrm{s}^{-1}$, i.e., one tenth of that in Figs. 2b and 3b. (d) Divergence difference field as in Fig. 3c except that the PV inversion is third order direct inversion. See text, section 6.

FIG. 4. Primitive-equation model integration at resolution T63, starting from the fields shown in Fig. 2. PV fields, Q, are shown every 2 days, with days numbered from the time corresponding to Fig. 2. The contour interval is the same as in Fig. 2c, i.e., $3{\times}10^{-8}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$, with the shading in the contour plot highlighting values lying between 4 and 6 of these units.

FIG. 5. Primitive-equation model integration at resolution T63, re-initialized by starting from the fields produced by third order direct inversion of the PV field shown in Fig. 2c. PV fields, Q, are shown at days 6 and 10. The contour interval is the same as in Figs. 2c and 4, i.e., $3{\times}10^{-8}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$, with the shading in the contour plot highlighting values lying between 4 and 6 of these units.

FIG. 6. Cumulative accuracy of direct PV inversion at different orders, tested using PV-conserving balanced model integrations at resolution T63 starting from the PV field shown in Fig. 2c. PV fields, Q, are shown at days 6 and 10 and are to be compared with those in Fig. 5. The contour interval is the same as in Fig. 5, i.e., $3{\times}10^{-8}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$, with the shading in the contour plot highlighting values lying between 4 and 6 of these units. (a),(b) Balanced model based on first order direct inversion. (c),(d) Balanced model based on second order direct inversion. (e),(f) Balanced model based on third order direct inversion.

FIG. 7. Cumulative accuracy of normal mode PV inversion at different orders, tested using PV-conserving balanced model integrations at resolution T63 starting from the PV field shown in Fig. 2c. PV fields, Q, are shown at days 6 and 10 and are to be compared with those in Fig. 5. The contour interval is the same as in Fig. 5, i.e., $3\times10^{-8}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$, with the shading in the contour plot highlighting values lying between 4 and 6 of these units. (a),(b) Balanced model based on first order normal mode inversion. (c),(d) Balanced model based on second order normal mode inversion. (e),(f) Balanced model based on third order normal mode inversion.

FIG. 8. (a) Departures from the behavior of the re-initialized primitive-equation integration corresponding to Fig. 5, measured by the r.m.s. differences, as functions of time, between the PV fields of that integration and the PV fields of the first order direct balanced model (D1, continuous curve), of the second order direct balanced model (D2, closely spaced dots), of the third order direct balanced model (D3, widely spaced dots), of the first order normal mode balanced model (NM1, widely spaced dashes), of the second order normal mode balanced model (NM2, short paired dashes), of the third order normal mode balanced model (NM3, dot-dashes), and primitive-equation integration corresponding to Fig. 4, not re-initialized (labeled PE, closely spaced dashes). (b) The same except that r.m.s. geopotential differences are used instead of r.m.s. PV differences. The initial imbalance revealed by the PE curves explains the similarity between Figs. 3b,c.

FIG. 9. (a) Divergence difference defined as δ from the primitive-equation model minus δ from a third order direct inversion performed in a new reference frame, rotating with angular velocity $\Omega = 7.8 \times 10^{-5}\mathrm{s}^{-1} > \Omega_{\mathrm{E}} = 7.3 \times 10^{-5}\mathrm{s}^{-1}$. The contour interval is $4 \times 10^{-9}\ \mathrm{s}^{-1}$. (b) Same as for Fig. 9a except that the inversion is performed in a reference frame rotating with $\Omega = 6.8 \times 10^{-5}\mathrm{s}^{-1} < \Omega_{\mathrm{E}}$.

FIG. 10. The r.m.s. synoptic-scale divergence error in the inversion of the PV shown in Fig. 2c as a function of reference-frame angular velocity Ω. Open triangles: first order normal mode inversion; open circles: second order normal mode inversion; open squares: third order normal mode inversion; solid circles: second order direct inversion; solid squares: third order direct inversion.

FIG. 11. Three different r.m.s. diagnostic error estimates (see text) for the third order direct inversion of the PV shown in Fig. 2c as a function of reference-frame angular velocity Ω. Triangles: estimate of $\partial^2\delta/\partial t^2$ obtained from the second time difference of δ calculated by timestepping the *balanced model* forward and backward by one timestep; circles: estimate of $\partial^2\delta/\partial t^2$ similarly obtained from the first time difference of δ_1; squares: estimate of $\partial^3\delta/\partial t^3$ similarly obtained from the second time difference of δ_1, multiplied by 10^3s. Such time differences are internal measures of inconsistency with primitive-equation evolution, requiring no reference to a parallel primitive-equation integration as in Fig. 10.